Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Statistical Planning and Inference 136 (2006) 3052–3069
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Fitting time series models for longitudinal survey data under informative sampling

Abdulhakeem A.H. Eideh[a,*], Gad Nathan[b]

[a]Department of Mathematics, Faculty of Science and Technology, Alquds University, Abu-Dies Campus, P.O. Box 20002, East Jerusalem, Palestine
[b]Department of Statistics, Hebrew University of Jerusalem, Israel

Received 21 November 2003; accepted 8 November 2004
Available online 6 April 2005

Abstract

The purpose of this paper is to account for informative sampling in fitting time series models, and in particular an autoregressive model of order one, for longitudinal survey data. The idea behind the proposed approach is to extract the model holding for the sample data as a function of the model in the population and the first-order inclusion probabilities, and then fit the sample model using maximum-likelihood, pseudo-maximum-likelihood and estimating equations methods. A new test for sampling ignorability is proposed based on the Kullback–Leibler information measure. Also, we investigate the issue of the sensitivity of the sample model to incorrect specification of the conditional expectations of the sample inclusion probabilities. The simulation study carried out shows that the sample-likelihood-based method produces better estimators than the pseudo-maximum-likelihood method, and that sensitivity to departures from the assumed model is low. Also, we find that both the conventional t-statistic and the Kullback–Leibler information statistic for testing of sampling ignorability perform well under both informative and noninformative sampling designs.

MSC: 62D05

Keywords: Autoregressive model; Informative sampling; Kullback–Leibler information measure; Pseudo-likelihood estimator; Sample-likelihood function

* Corresponding author. Tel.: +972 2 2219555; fax: +972 2 2796960.
E-mail address: msabdul@palnet.com (A.A.H. Eideh).

0378-3758/$ - see front matter
doi:10.1016/j.jspi.2004.12.003

1. Introduction

Each year around the world statistical agencies and other organizations conduct thousands of sample surveys to obtain information required for decision-making or policymaking. Although many of the surveys are cross-sectional, there is increasing interest in longitudinal surveys, for which similar measurements on the same variables or characteristics are made on the same units at different points of time (occasions or waves). Each series of observations for the same unit can be viewed as a time series, usually of short length. Analyzing the measurements for all the sampled subjects permits the fitting of a low-order time series models, despite the short lengths of the individual series. There are several reasons for the increased prevalence of longitudinal surveys, which often have different objectives—see for example Binder (1998), Binder and Hidiroglou (1988), Duncan and Kalton (1987), and Smith (1978). In recent years, there has been a growing interest in analyzing data collected from longitudinal surveys that use complex sampling designs, as is often the case in efficient survey sampling. This interest reflects the expansion in requirements by policy makers and social scientists for in-depth studies of social processes over time, rather than of one-time "snap-shots" provided by cross-sectional analyses. Examples of longitudinal surveys, some of which are based on complex sample designs, and of the issues involved in their design and analysis can be found in Friedlander et al. (2002), Herriot and Kasprzyk (1984), and Nathan (1999). In many of the cases described in these papers, samples are selected for the first round and continue to serve for several rounds.

Most of the previous work in the area of the analysis of longitudinal data deals with it in a nonsurvey context; see for example, Diggle et al. (1994), Lawless (2003), and Zeger et al. (1988). In the survey context, see Feder et al. (2000) and Skinner and Holmes (2003).

Standard analysis of longitudinal survey data often fails to account for the complex nature of the sampling design, such as the use of unequal selection probabilities, clustering, post-stratification and other kinds of weighting used for the treatment of nonresponse. Thus it does not incorporate all the design variables in the analysis model, either because there may be too many of them or because they are not of substantive interest. However, if the sampling design is informative, in the sense that the outcome variable is correlated with design variables not included in the model, even after conditioning on the model covariates, standard estimates of the model parameters can be severely biased, leading possibly to false inference—see for example Pfeffermann (1993,1996), Hoem (1989), Fienberg (1989), Kasprzyk et al. (1989), Skinner (1994) and Skinner et al. (1989).

It should be emphasized that standard inference may be biased even when the original sample is a simple random sample, due to nonresponse, attrition and imperfect frames that results in de facto a posterior differential inclusion probabilities. Special features of longitudinal studies, such as late additions of individuals who join panel households, can also lead to de facto unequal inclusion probabilities. In this paper we assume full response. The problem of nonresponse, attrition or dropout in longitudinal surveys is discussed, for instance, by Diggle and Kenward (1994), Pfeffermann and Nathan (2001) and Miller et al. (2001).

The aim of this paper is to fit time series models and in particular an autoregressive model of order one for longitudinal survey data, when the sampling design is informative. In Section 2, we introduce the marginal distribution of longitudinal sample observations under informative sampling. Section 3 discusses the population and sample autoregressive model of order one for longitudinal surveys under informative sampling. Section 4 presents methods of estimation and variance estimation. Section 5 is devoted to testing sampling ignorability and Section 6 presents the simulation study with discussion.

2. Sample distribution of longitudinal observations

Some recent work has considered the definition of the sample distribution under informative sampling for cross-sectional data. Sample survey data may be viewed as the outcome of two processes: the process that generates the values of units in the finite population, often referred to as the superpopulation model, and the process of selecting the sample units from the finite population, known as the sample selection mechanism. Analytic inference from repeated survey data refers to the superpopulation model. When the sample selection probabilities depend on the values of the model response variable, even after conditioning on auxiliary variables, the sampling mechanism becomes informative and the selection effects need to be accounted for in the inference process.

Pfeffermann et al. (1998) propose a general method of inference on the population (model) distribution under informative sampling that consists of approximating the parametric distribution of the sample measurements for given population distributions and first-order sample selection probabilities. The sample distribution is defined as the distribution of the sample measurements, given these measurements are sampled. Under informative sampling, this distribution is different from the corresponding population distribution. In this section, we extend these results to study the case of longitudinal panel sample observations under informative sampling.

Consider a finite population, U, of size N, assumed to remain constant over time. The individuals of U are potentially measured repeatedly over time. Let y_{it} be the value of the response variable y, attached to the individual i at time t, $(t = 1, \ldots, T;\ i = 1, \ldots, N)$. These values are assumed to be related to the previous values $\{y_{it'}, t' = 1, \ldots, t-1\}$ of the same individual under some correlation structure. Let $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ be the vector of T responses on the ith individual, $i = 1, \ldots, N$. Assume that the values $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are independent realizations with pdf $f_{\mathrm{p}}(\mathbf{y}_i|\boldsymbol{\theta})$, which depends on the parameter $\boldsymbol{\theta}$. Since the elements of $\mathbf{y}_i$ are dependent measurements on a single subject that constitute a time series, the population pdf of $\mathbf{y}_i$ can be written as

$$f_{\mathrm{p}}(\mathbf{y}_i|\boldsymbol{\theta}) = f_{\mathrm{p}}(y_{i1}|\boldsymbol{\theta}) \prod_{t=2}^{T} f_{\mathrm{p}}(y_{it}|H_{it-1};\boldsymbol{\theta}) = f_{\mathrm{p}}(y_{i1}|\boldsymbol{\theta}) f_{\mathrm{p}}(y_{i2}, \ldots, y_{iT}|y_{i1};\boldsymbol{\theta}), \quad (1)$$

where H_{it-1} denotes the set of measurements on the ith individual up to and including y_{it-1}.

In what follows, let $\mathbf{z}=\{z_1, \ldots, z_N\}$ be the values of a known design variable employed for the sample selection process and consider the inclusion probabilities $\pi_i = \Pr(i \in s|\mathbf{y}_i, \mathbf{z}_i) = \Pr(I_i = 1|\mathbf{y}_i, \mathbf{z}_i)$ as realizations of random variables with conditional probability density function $f_{\mathrm{p}}(\pi_i|y_i, \gamma)$, $i=1, \ldots, N$, where $I_i=1$ if unit i is in the sample and $I_i=0$ otherwise. We consider the case where the sample is a panel sample selected at time $t=1$ and all units remain in the sample till time $t = T$. Then it is intuitively reasonable to assume that the first-order inclusion probabilities depend on the population values of the response variable at the first occasion only, the values y_{i1}, and the values of the design variable used for the sample selection, but not included in the working model under consideration. Then we have the following results.

Lemma 1 (*Sample distribution of longitudinal observations*). *Let $f_p(\mathbf{y}_i|\boldsymbol{\theta})$ be the population distribution of $\mathbf{y}_i$.*

1. *The sample distribution of $\mathbf{y}_i$ is given by*

$$f_s(\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \frac{E_p(\pi_i|\mathbf{y}_i; \boldsymbol{\gamma})}{E_p(\pi_i|\boldsymbol{\theta}, \boldsymbol{\gamma})} f_p(\mathbf{y}_i|\boldsymbol{\theta}). \tag{2}$$

2. *If we assume that π_i depends only on y_{i1} and $\mathbf{z}_i$, then the sample distribution of $\mathbf{y}_i$ is given by*

$$f_s(\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\gamma}) = f_s(y_{i1}|\boldsymbol{\theta}, \boldsymbol{\gamma}) \prod_{t=2}^{T} f_p(y_{it}|H_{it-1}; \boldsymbol{\theta}), \tag{3}$$

where

$$f_s(y_{i1}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = \frac{E_p(\pi_i|y_{i1}; \boldsymbol{\gamma})}{E_p(\pi_i|\boldsymbol{\theta}, \boldsymbol{\gamma})} f_p(y_{i1}|\boldsymbol{\theta})$$

is the sample distribution of y_{i1}.

Proof. Application of Eq. (3.4) of Pfeffermann et al. (1998).

Comment 1: Note that $E_p(\pi_i|\mathbf{y}_i) = E_{\mathbf{z}_i|\mathbf{y}_i} E_p(\pi_i|\mathbf{y}_i, \mathbf{z}_i)$, so that $\mathbf{z}_i$ is integrated out in (2).

Comment 2: Note that the sample and the population distribution of longitudinal observations are, in general different. In particular, in (3) the difference between the population and sample distributions is due to the difference in the distribution of y_{i1} only. The product term in (3) does not change. This is because the selected sample is a panel sample selected at time $t = 1$ and all units remain in the sample till time $t = T$.

Comment 3: According to (3), we can see that, for a given population distribution, the sample distribution is completely determined by the specification of the population conditional expectations of sample inclusion probabilities, $E_p(\pi_i|y_{i1}; \boldsymbol{\gamma})$.

Pfeffermann et al. (1998) introduced two alternative approximation models for this population conditional expectation; see also Skinner (1994):

(a) *Exponential inclusion probability model*:

$$E_p(\pi_i|y_{i1}; a_0, a_1) = \exp(a_0 + a_1 y_{i1}). \tag{4}$$

In this case, we can show that the sample distribution of $\mathbf{y}_i$ is given by

$$f_s(\mathbf{y}_i|\boldsymbol{\theta}, a_1) = \frac{\exp(a_1 y_{i1}) f_p(y_{i1}|\boldsymbol{\theta})}{M_p(a_1, \boldsymbol{\theta})} \prod_{t=2}^{T} f_p(y_{it}|H_{it-1}; \boldsymbol{\theta}), \tag{5}$$

where a_0 and a_1 are informativeness parameters to be estimated from the sample and $M_{\mathrm{p}}(a_1, \boldsymbol{\theta}) = E_{\mathrm{p}}[\exp(a_1 y_{i1})]$ is the moment generating function of the population distribution of y_{i1}.

(b) *Linear inclusion probability model*:

$$E_{\mathrm{p}}(\pi_i | y_{i1}; b_0, b_1) = (b_0 + b_1 y_{i1}). \tag{6}$$

In this case the sample distribution of $\mathbf{y}_i$ is given by

$$f_{\mathrm{s}}(\mathbf{y}_i | \boldsymbol{\theta}, b_0, b_1) = \frac{(b_0 + b_1 y_{i1}) f_{\mathrm{p}}(y_{i1} | \boldsymbol{\theta})}{[b_0 + b_1 E_{\mathrm{p}}(y_{i1})]} \prod_{t=2}^{T} f_{\mathrm{p}}(y_{it} | H_{it-1}; \boldsymbol{\theta}), \tag{7}$$

where b_0 and b_1 are informativeness parameters to be estimated from the sample.

Pfeffermann et al. (1998) proved that, under certain regularity conditions and for the majority of sampling schemes that are used in practice, such as probability proportional to size with or without replacement, successive sampling, rejective sampling, and Rao–Sampford's method; if s consists of n distinct units and if the population measurements are independent, then as $N \to \infty$ (with n fixed) the sample responses are asymptotically independent. Thus, we can apply standard inference procedures to complex survey data by using the marginal sample distribution for each unit in the sample.

3. Autoregressive model of order one for longitudinal survey data under informative sampling

3.1. Population model

Let the observed measurements, y_{it}, be a realization of random variables which follow the first-order autoregressive (AR) model; that is

$$y_{it} - \mu = \phi(y_{it-1} - \mu) + \varepsilon_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{8}$$

where the errors $\{\varepsilon_{it}\}$ are independent and normally distributed, with zero mean and variance σ^2, and $|\phi| < 1$. Since $\varepsilon_{it} \underset{\mathrm{p}}{\sim} N(0, \sigma^2)$ and $y_{it} | H_{it-1} \underset{\mathrm{p}}{\sim} N\{\phi(y_{it-1} - \mu) + \mu, \sigma^2\}$, then the population pdf of the ith response $\mathbf{y}_i = (y_{i1}, \ldots, y_{iT})'$ is

$$f_{\mathrm{p}}(\mathbf{y}_i) = f_{\mathrm{p}}(y_{i1})(2\pi\sigma^2)^{[(T-1)/2]} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=2}^{T} [y_{it} - \mu - \phi(y_{it-1} - \mu)]^2\right\}. \tag{9}$$

If y_{i1} is regarded as fixed in repeated realizations, then $f_{\mathrm{p}}(y_{i1})$ can be dropped because it does not affect the inference, in such cases there is no problem of informativeness. However, when y_{i1}, $i = 1, \ldots, N$ are considered as random variables, then, under the stationarity condition of AR (1), it follows that, $y_{i1} \underset{\mathrm{p}}{\sim} N(\mu, \sigma^2/[1 - \phi^2])$.

According to this interpretation the population pdf of $\mathbf{y}_i$ is

$$f_{\rm p}(\mathbf{y}_i) = c(\sigma^2, \phi) \exp\left(-\frac{1}{2\sigma^2}\left[(1-\phi^2)(y_{i1}-\mu)^2 + \sum_{t=2}^{T}\{(y_{it}-\mu) - \phi(y_{it-1}-\mu)\}^2\right]\right), \tag{10}$$

where $c(\sigma^2, \phi) = (2\pi\sigma^2)^{-T/2}(1-\phi^2)^{1/2}$.

3.2. Sample distribution under the AR (1) model

(a) *Exponential inclusion probability model*: $E_{\rm p}(\pi_i|y_{i1}) = \exp(a_0 + a_1 y_{i1})$: According to (5) we can show that, under this model, the sample log-likelihood function is given by

$$\begin{aligned} l_{\rm e}(\mu, \phi, \sigma^2, a_1) = & -\frac{nT}{2}\log(\sigma^2) + \frac{n}{2}\log(1-\phi^2) \\ & - \frac{1-\phi^2}{2\sigma^2}\sum_{i\in s}\left(y_{i1} - \mu - \frac{a_1\sigma^2}{1-\phi^2}\right)^2 \\ & + \sum_{i\in s}\left[-\frac{1}{2\sigma^2}\sum_{t=2}^{T}\{y_{it} - \mu - \phi(y_{it-1}-\mu)\}^2\right]. \end{aligned} \tag{11}$$

Thus the sample pdf belongs to the same family as the population pdf but differs only in the mean of the first variable y_{i1}. The sample pdf of y_{i1} is also normal but the mean μ in the population pdf changes to $\mu + (a_1\sigma^2/1-\phi^2)$, for the sample distribution. The contribution of the component $\prod_{t=2}^{T} f_{\rm p}(y_{it}|H_{it-1})$ does not change.

(b) *Linear inclusion probability model*: $E_{\rm p}(\pi_i|y_{i1}) = b_0 + b_1 y_{i1}$: Under this approximation, using (7), the sample log-likelihood function is

$$\begin{aligned} l_{\rm l}(\mu, \phi, \sigma^2, b_0, b_1) = & -n\log(b_0 + b_1\mu) - \frac{nT}{2}\log(\sigma^2) \\ & + \frac{n}{2}\log(1-\phi^2) - \frac{1-\phi^2}{2\sigma^2}\sum_{i\in s}(y_{i1}-\mu)^2 \\ & + \sum_{i\in s}\left[-\frac{1}{2\sigma^2}\sum_{t=2}^{T}\{y_{it} - \mu - \phi(y_{it-1}-\mu)\}^2\right]. \end{aligned} \tag{12}$$

4. Estimation

In this section, we extend the methods of estimation for cross-sectional survey data to longitudinal survey data, in particular the method based on the sample distribution function (Pfeffermann et al., 1998; Sverchkov and Pfeffermann, 1999; Pfeffermann and Sverchkov, 2003) and the pseudo-maximum-likelihood method (Binder, 1983; Skinner, 1989).

4.1. Two-step estimation

In our case the parameter on which the population pdf depends is $\boldsymbol{\theta}=(\mu,\ \phi,\ \sigma^2)$, whereas the parameters on which the sample pdf depends are $\boldsymbol{\theta}$ and a_1, see Eq. (11) or $\boldsymbol{\theta}$, b_0 and b_1, see Eq. (12). Thus the parameters of the sample distribution include the parameters of the population distribution. Hence, the parameters of the population distribution can be estimated using the sample distribution, for instance, by the following two-step method. As pointed out by Pfeffermann et al. (1998), the use of this two-step procedure becomes necessary when the conditional expectation of the first-order sample selection probabilities is exponential, since in this case there is a problem of identifiability.

In practice, the conditional expectations of the sample inclusion probabilities, $E_{\mathrm{p}}(\pi_i|y_{i1})$, are not known and usually the only data available to the analyst for the first time period are $\{y_{i1}, w_i;\ i \in s\}$, where $w_i = 1/\pi_i$ are the sample weights. The estimation of $E_{\mathrm{p}}(\pi_i|y_{i1})$, using only the sample data can be based on the following relationship, due to Pfeffermann and Sverchkov (1999)

$$E_{\mathrm{s}}(w_i|y_{i1}) = \frac{1}{E_{\mathrm{p}}(\pi_i|y_{i1})}. \tag{13}$$

The prominent feature of this relationship is that the expectation of the population conditional sample inclusion probabilities can be identified and estimated from the sample data.

Thus for the exponential inclusion probability model, two-step estimation proceeds as follows:

Step 1: *Estimation of* a_0 *and* a_1: According to (13) we estimate a_0 and a_1 by regressing $-\log(w_i)$ against $y_{i1}, i \in s$.

Step 2: Substituting the ordinary least-squares estimator, $\tilde{a}_1$ of a_1 in the sample log-likelihood (11), estimates of the remaining parameters are obtained by differentiating the sample log-likelihood, $l_{\mathrm{e}}(\mu, \varphi, \sigma^2, \tilde{a}_1)$, with respect to each of the components of $\boldsymbol{\theta} = (\mu, \phi, \sigma^2)$. The resulting equations are not linear in the parameters, so a numerical solution is needed to compute the ML estimates.

For the linear inclusion probability model given in (6), we first estimate the informativeness parameters using (13), by regressing $1/w_i$ against $y_{i1},\ i \in s$, and then substituting the resulting estimates, $\tilde{b}_0$ and $\tilde{b}_1$ in (12). The ML estimators of the parameters, $\boldsymbol{\theta}$, can be obtained by differentiating the resulting sample log-likelihood function, $l_1(\mu, \phi, \sigma^2, \tilde{b}_0, \tilde{b}_1)$, with respect to each of the components of $\boldsymbol{\theta} = (\mu, \phi, \sigma^2)$.

4.2. Pseudo-likelihood approach

The census maximum-likelihood estimator of $\boldsymbol{\theta}$ solves the census likelihood equations, which in our case are

$$U(\boldsymbol{\theta}) = \sum_{i=1}^{N} \frac{\partial}{\partial \boldsymbol{\theta}} \left\{ -\frac{T}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}(y_{i1}-\mu)^2(1-\phi^2) \right\} - \frac{1}{2\sigma^2}\sum_{i=1}^{N}\sum_{t=2}^{T}\frac{\partial}{\partial \boldsymbol{\theta}}\{(y_{it}-\mu)-\phi(y_{it-1}-\mu)\}^2 = 0. \tag{14}$$

The pseudo-maximal-likelihood (PML) estimator is defined as the solution of $\hat{U}(\boldsymbol{\theta}) = 0$ where $\hat{U}(\boldsymbol{\theta})$ is a sample estimator of the census log-likelihood, $U(\boldsymbol{\theta})$. See Binder (1983).

We consider two possibilities for $\hat{U}(\boldsymbol{\theta})$

$$\hat{U}_w(\boldsymbol{\theta}) = \sum_{i\in s} w_i \frac{\partial}{\partial\boldsymbol{\theta}}\left\{-\frac{T}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}(y_{i1}-\mu)^2(1-\phi^2)\right\} - \frac{1}{2\sigma^2}\sum_{i\in s}\sum_{t=2}^{T} w_i \frac{\partial}{\partial\boldsymbol{\theta}}\{(y_{it}-\mu)-\phi(y_{it-1}-\mu)\}^2, \tag{15}$$

or

$$\hat{U}_{ws}(\boldsymbol{\theta}) = \sum_{i\in s} w_i \frac{\partial}{\partial\boldsymbol{\theta}}\left\{-\frac{T}{2}\log(\sigma^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma^2}(y_{i1}-\mu)^2(1-\phi^2)\right\} - \frac{1}{2\sigma^2}\sum_{i\in s}\sum_{t=2}^{T} \frac{N}{n} \frac{\partial}{\partial\boldsymbol{\theta}}\{(y_{it}-\mu)-\phi(y_{it-1}-\mu)\}^2. \tag{16}$$

The basic idea behind the use of self-weighting in (16) for the time periods $t \geqslant 2$ is that: for longitudinal survey data under informative sampling, the model holding for the sample units in the first occasion is different from the model holding in the population for that time period, but the conditional distribution for the remaining time periods does not change. Strictly speaking, this is no longer a pseudo-likelihood estimator and is similar to the estimating equation approach based on relationships between sample and population distributions suggested in Pfeffermann and Sverchkov (2003). The estimators are obtained by solving the equations derived from (15) or (16). The log-likelihood equations again are not linear in the parameters, so a numerical optimization is needed to compute the estimates.

If y_{i1} is regarded as fixed, the probability-weighted estimator is the solution of the equation

$$\frac{\partial \hat{l}_c(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}} = \frac{\partial}{\partial\boldsymbol{\theta}}\left\{-\frac{n(T-1)}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i\in s}\sum_{t=2}^{T} w_i[(y_{it}-\mu)-\phi(y_{it-1}-\mu)]^2\right\} = 0. \tag{17}$$

4.3. Variance estimation

For the variance estimation of $\hat{\boldsymbol{\theta}}$, we consider the use of the inverse of the Fisher information matrix, following Pfeffermann and Sverchkov (1999, 2003). We first consider estimating the conditional variance of $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi}, \hat{\sigma}^2)$, given that the informativeness parameters a_1, b_0 and b_1 are held fixed at their estimated values. The conditional Fisher information matrix evaluated at $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi}, \hat{\sigma}^2)$ is given by

$$\hat{V}_{\text{sample}}(\hat{\boldsymbol{\theta}}) = [I_{\text{sample}}(\hat{\boldsymbol{\theta}})]^{-1} = \left\{-\frac{1}{n}\left[\frac{\partial^2 l(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'\partial\boldsymbol{\theta}}\right]\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right\}^{-1}. \tag{18}$$

For instance, under the linear inclusion probability model; the entries of $I_{\text{sample}}(\boldsymbol{\theta})$ are given by

$$\frac{\partial^2 l_1}{\partial\mu^2} = -\frac{n}{\sigma^2}\left[(T-1)(1-\phi)^2 + (1-\phi^2) - \frac{\tilde{b}_1^2\sigma^2}{\tilde{b}_0 + \tilde{b}_1\mu}\right],$$

$$\frac{\partial^2 l_1}{\partial\phi\partial\mu} = -\frac{1}{\sigma^2}\sum_{i\in s}\left\{Y_i^* + (1-\phi)\sum_{t=2}^{T}(y_{it-1}-\mu) + 2\phi(y_{i1}-\mu)\right\},$$

$$\frac{\partial^2 l_1}{\partial\sigma^2\partial\mu} = -\frac{1}{\sigma^4}\sum_{i\in s}\{Y_i^*(1-\phi) + (1-\phi^2)(y_{i1}-\mu)\},$$

$$\frac{\partial^2 l_1}{\partial\phi^2} = -\frac{1}{\sigma^2}\sum_{i\in s}\left[\sum_{t=2}^{T}(y_{it-1}-\mu)^2 + \sigma^2\frac{1+\phi^2}{(1-\phi^2)^2} - (y_{i1}-\mu)^2\right],$$

$$\frac{\partial^2 l_1}{\partial\sigma^2\partial\phi} = -\frac{1}{\sigma^4}\sum_{i\in s}\left\{\sum_{t=2}^{T}[(y_{it}-\mu) - \phi(y_{it-1}-\mu)](y_{it-1}-\mu) + \phi(y_{i1}-\mu)^2\right\},$$

$$\frac{\partial^2 l_1}{\partial\sigma^4} = -\frac{1}{\sigma^4}\sum_{i\in s}\left\{-\frac{T}{2} + \frac{1}{\sigma^2}\sum_{t=2}^{T}[(y_{it}-\mu) - \phi(y_{it-1}-\mu)]^2 + \frac{1-\phi^2}{\sigma^2}(y_{i1}-\mu)^2\right\},$$

where

$$Y_i^* = \sum_{t=2}^{T}[(y_{it}-\mu) - \phi(y_{it-1}-\mu)].$$

Similar results are obtained under the exponential inclusion probability model.

In order to estimate the unconditional variance the unconditional sample likelihoods must be used—see Eqs. (11) and (12). For the linear case the parameters are identifiable, so that the MLE of $\mu, \phi, \sigma^2, b_0, b_1$ can be obtained directly without using the two-step method and explicit expressions for the Fisher information matrix are available. For the exponential case explicit expressions are not available, since the parameters are not identifiable.

For such cases, an alternative to the Fisher information method is the bootstrap approach for variance estimation. This is well founded under informative sampling, since under the majority of sampling schemes used in practice the sample measurements are asymptotically independent with respect to the sample distribution. Let $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi}, \hat{\sigma}^2)$ be the sample ML of $\boldsymbol{\theta} = (\mu, \phi, \sigma^2)$ based on Eqs. (11) or (12) and $\hat{\boldsymbol{\theta}}_b = (\hat{\mu}_b, \hat{\phi}_b, \hat{\sigma}_b^2)$ be the ML estimator computed from the bootstrap sample $b = 1, \ldots, B$, with the same sample size, drawn by simple random sampling with replacement from the original sample—the sample drawn under informative sampling design. The bootstrap variance estimator of $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\phi}, \hat{\sigma}^2)$ is defined as

$$\hat{V}_{\text{boot}}(\hat{\boldsymbol{\theta}}) = \frac{1}{B}\sum_{b=1}^{B}(\hat{\boldsymbol{\theta}}_b - \bar{\hat{\boldsymbol{\theta}}}_{\text{boot}})(\hat{\boldsymbol{\theta}}_b - \bar{\hat{\boldsymbol{\theta}}}_{\text{boot}})', \quad \text{where } \bar{\hat{\boldsymbol{\theta}}}_{\text{boot}} = \frac{1}{B}\sum_{b}^{B}\hat{\boldsymbol{\theta}}_b.$$

As pointed by Pfeffermann and Sverchkov (2003), a possible advantage of the use of the bootstrap variance estimator in the present context is that it accounts for all sources of variation, including that due to the estimation of the unknown informativeness parameters, so that it estimates the unconditional variance.

5. Testing for sampling ignorability

A natural question arising for informative sampling is how to test if the sampling design can be ignored for the inference process, given the available design information. Under the assumptions of Section 2, it is easy to verify that $f_s(\mathbf{y}_i|\boldsymbol{\theta}, \boldsymbol{\gamma}) = f_p(\mathbf{y}_i|\boldsymbol{\theta})$, that is the sampling design is ignorable (noninformative), if and only if

$$f_s(y_{i1}|\boldsymbol{\theta}, \boldsymbol{\gamma}) = f_p(y_{i1}|\boldsymbol{\theta}). \tag{19}$$

5.1. Conventional correlation t-statistic

Condition (19) allows us to use the t-correlation test for testing sampling ignorability; see Pfeffermann and Sverchkov (1999). If we have no auxiliary variables, then we can show that the test of sampling ignorability is equivalent to testing the set of hypotheses:

$$\mathrm{H}_{0k} = \mathrm{Corr}_s(y_{i1}^k, w_i) = 0, \quad k = 1, 2, \ldots, \tag{20}$$

where Corr_s is the correlation under the sample distribution.

The conventional t-statistic can be used to test this set of hypotheses. This test requires that the all moments of the distribution exist.

5.2. Kullback–Leibler information test

A new test for sampling ignorability we propose is based on the Kullback–Leibler (K–L) information measure; see Kullback (1978). For instance, under the exponential inclusion probability model, the condition (19) implies that the test of sampling ignorability is equivalent to testing the null hypothesis

$$\mathrm{H}_0 : f_p(y_{i1}|\boldsymbol{\theta}) = f_s(y_{i1}|\boldsymbol{\theta}, a_1) \quad \text{or } a_1 = 0, \tag{21a}$$

against the alternative hypothesis

$$\mathrm{H}_1 : f_p(y_{i1}|\boldsymbol{\theta}) \neq f_s(y_{i1}|\boldsymbol{\theta}, a_1) \quad \text{or } a_1 \neq 0. \tag{21b}$$

We can show that the minimum discrimination information (for a single observation) from the sample log-likelihood given by (11) is

$$I(f_s : f_p) = E_s\left[\log \frac{f_s(y_{i1}|\boldsymbol{\theta}, a_1)}{f_p(y_{i1}|\boldsymbol{\theta})}\right] = \frac{a_1}{2}\left(\frac{a_1\sigma^2}{1-\phi^2}\right). \tag{22}$$

Note that $I(f_s : f_p)$ is a product of $a_1/2$ and the amount of change in the location parameter, and is equal zero if and only if $f_s(y_{i1}) = f_p(y_{i1})$, that is, if $a_1 = 0$.

Now let s denote the random sample $y_{11}, \ldots, y_{n1}$ of n independent and identically distributed observations. Then the estimate of the minimum discrimination information given in (22) is

$$\hat{I}(f_s : f_p) = \hat{I}(\mathrm{H}_1 : \mathrm{H}_0) = n\frac{\tilde{a}_1^2\hat{\sigma}^2}{2(1-\hat{\phi}^2)}, \tag{23}$$

where $\tilde{a}_1$ and $\hat{\boldsymbol{\theta}}$ are the appropriate estimators of a_1 and $\boldsymbol{\theta}$, respectively. Asymptotically, under certain regularity conditions and under the null hypothesis H_0, $2\hat{I}(\mathrm{H}_1 : \mathrm{H}_0)$ given in (23) has an asymptotic chi-square distribution with one degree of freedom, see Kullback (1978, Section 5.5). Thus, asymptotically

$$\Pr(2\hat{I}(\mathrm{H}_1 : \mathrm{H}_0) \geqslant \chi^2_{2\alpha,1}) = \alpha. \tag{24}$$

In developing our maximum-likelihood estimators and the K–L information test for ignorability of sampling design for the AR (1) model, we have derived the estimators and the test of informativeness assuming a correctly specified model. In practice, one is often obliged to specify the model at the same time that one is constructing the estimates and the information test. A number of criteria have been suggested for use in model selection. For discussions of model evaluation and selection in time series analysis, see Fuller (1996, pp. 438–439) and Harvey (1993, pp. 73–80). However in our case, the sample-likelihood functions given in (11) and (12) are different from the classical likelihood function, where selection bias is absent. As mentioned before, since the ML estimators defined by (11) and (12) are only for the population parameter $\boldsymbol{\theta} = (\mu, \phi, \sigma^2)$, with the estimates of the informativeness parameters a_1, b_0 and b_1 held fixed at their estimated values, then well-known methods can be applied. In particular we can use the AIC criterion introduced by Akaike (1973), defined by

$$\mathrm{AIC} = -2l(\hat{\boldsymbol{\theta}}) + 2k = -2l(\hat{\mu}, \hat{\phi}, \hat{\sigma}^2) + 2k,$$

where $l(\hat{\boldsymbol{\theta}}) = l(\hat{\mu}, \hat{\phi}, \hat{\sigma}^2)$ is the sample log-likelihood given by (11) or (12) and k is the dimension of $\boldsymbol{\theta}$, which is equal to 3 in our case. We can fit several alternative models and choose the one with the smallest AIC.

6. Simulation study with discussion

In order to assess the performance of the estimators obtained using parametric and pseudo-likelihood methods and compare them with the classical estimators obtained under simple random sampling, we designed and executed a simulation study.

6.1. Generation of population values

The population values were generated in 3 steps:

Step 1: We generated independently univariate normal values of y_{i1} at the first time period $(t = 1)$ from: $y_1 \sim N(\mu, \sigma^2/(1-\phi^2))$, where $\mu = 2$, $\phi = 0.8$ and $\sigma = 0.5$.

Step 2: We generated independently the population values of the random error term from: $\varepsilon_{it} \sim N\{0, (0.5)^2\}$, $i = 1, \ldots, 5000$; $t = 1, \ldots, T$ with $T = 3$ and 10.

Step 3: The population values for the time periods $t = 2, 3, \ldots, T$ were generated using the AR (1) model as follows: $y_{it} = 2 + 0.8(y_{it-1} - 2) + \varepsilon_{it}$, $i = 1, \ldots, 5000$.

6.2. Sample selection

Samples of size $n=500$ were selected by probability proportional to size (PPS) systematic sampling, with the size variable, z_i, defined in three different ways:

I—*Exponential sampling*: $z_i(e) = \exp(0.5 + 0.2y_{i1} + u_i),\ u_i \sim U(0, 1)$.
II—*Linear sampling*: $z_i(l) = 4 + 5y_{i1} + u_i,\ u_i \sim U(0, 25)$.
III—*Ignorable sampling*: $z_i(\text{ig}) = \exp(u_i),\ u_i \sim U(0, 4)$.

Under exponential and linear sampling the sampling design is informative, while under ignorable sampling the sampling design is noninformative.

The first-order inclusion probabilities and the sampling weights were defined, respectively, by

$$\pi_i(j) = [nz_i(j)] \Bigg/ \sum_{i=1}^{N} z_i(j) \quad \text{and} \quad w_i(j) = 1/\pi_i(j), \quad j = e, l, \text{ or ig}$$

After generating the population values and specifying the sampling design, the population was simulated $R = 500$ times for two periods of length $T = 3, 10$ (results for $T = 10$ are not given, for lack of space) and for each simulated population three samples were independently drawn using PPS(e), PPS(l) and PPS(ig). Data from these samples were then used to estimate the informativeness parameters and then the population parameters using the different methods described in Section 4.

6.3. The estimators compared

The parameters estimated in our study are the mean μ, the autocorrelation ϕ, and the variance σ^2. We consider six different estimators, four of them based on the assumption that y_{i1} is a random variable and the remaining two estimators based on the assumption that y_{i1} is fixed. These estimators are described as follows:

UWML—Unweighted (exact) maximum likelihood (ML) for the case where the sampling design is ignorable—solutions to (14) with N replaced by n.

WSML—Weighted ML for $t = 1$ and self-weighted for $t \geqslant 2$. The estimator obtained by using (16).

PWML—Probability weighted estimator or pseudo-ML (15).

SMLE—The estimator obtained by maximizing the resulting sample log-likelihood function given by (11), based on the exponential inclusion probability model.

SMLL—The estimator obtained by maximizing the resulting sample log-likelihood given by (12), based on the linear inclusion probability model.

CUWML—Unweighted (exact) conditional ML obtained by solving (17) with $w_i = 1$ (y_{i1} regarded as fixed).

CWML—Weighted conditional ML obtained by solving (17) (y_{i1} regarded as fixed).

Comment: The likelihood functions are maximized using the *nlminb* function within S-Plus (Statistical Sciences, 1990).

The relative bias of $\hat{\theta}$ is estimated by

$$\mathrm{RB}(\hat{\theta}) = \frac{\mathrm{MSE}(\hat{\theta})}{\theta} = \frac{1}{\theta}\left\{\frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_i - \theta)^2\right\}$$

and the relative root-mean-square error of $\hat{\theta}$ is estimated by

$$\mathrm{RRMSE}(\hat{\theta}) = \frac{\mathrm{sqrt}[\mathrm{MSE}(\hat{\theta})]}{\theta} = \frac{1}{\theta}\left[\frac{1}{500}\sum_{i=1}^{500}(\hat{\theta}_i - \theta)^2\right]^{0.5}.$$

6.4. Sensitivity analysis

An important aspect of the use of the sample model for statistical inference is its sensitivity to wrong specification of the conditional expectation of the first-order sample inclusion probabilities, $E_{\mathrm{p}}(\pi_i | y_{i1}, \gamma)$.

To investigate this issue we assume that sample selection is based on the exponential sampling or on the linear sampling and then apply first the correct model and then the incorrect model for estimation. The results obtained are compared under the correct and incorrect models.

6.5. Results

The results of the simulation study are summarized in Table 1 as averages over the 500 samples selected under the three sampling schemes—the exponential, linear and ignorable (noninformative) sampling schemes and contains the results of the sensitivity analysis.

The main findings from Table 1 are as follows:

1. The unweighted ML estimators of the population mean, μ, have higher relative biases and relative root-mean-square errors for both of the informative sampling schemes—exponential and linear. These relative biases and relative root-mean-square errors are reduced substantially by the other estimation methods, WSML, PWML, SMLL and SMLE, under both the two informative sampling schemes.
2. The relative biases and relative root-mean-square errors of the estimators of all three parameters behave approximately the same under the four methods of estimation, WSML, PWML, SMLL and SMLE, and for the two informative sampling schemes, while under the ignorable case the SMLL and SMLE are more efficient than WSML and PWML.
3. The smallest relative bias occurs under the sample distribution method of estimation when the sampling design is exponential while under linear sampling this occurs for ϕ and σ^2 in most of the cases. Also the WSML and PWML have smallest biases for μ when the sampling is linear.

Table 1
Relative biases (RB—italics) and relative root-mean-square error (RRMSE—bold) of estimators under five estimation methods ($T = 3$ time periods)

True model	Estimation method					
	Parameter	UWML	WSML	PWML	SMLE	SMLL
Exponential	μ	*0.0573*	*0.0013*	*0.0008*	*−0.0009*	*−0.0032*
		0.0600	**0.0184**	**0.0185**	**0.0184**	**0.0188**
	ϕ	*−0.0002*	*−0.0018*	*−0.0013*	*0.0054*	*0.0050*
		0.0157	**0.0157**	**0.0161**	**0.0165**	**0.0165**
	σ^2	*0.0035*	*−0.0076*	*−0.0081*	*−0.0055*	*−0.0011*
		0.0422	**0.0428**	**0.0453**	**0.0423**	**0.0423**
Linear	μ	*0.0537*	*0.0020*	*0.0016*	*0.0056*	*0.0084*
		0.0562	**0.0182**	**0.0184**	**0.0185**	**0.0174**
	ϕ	*−0.0063*	*−0.0046*	*−0.0039*	*−0.0012*	*−0.0033*
		0.0185	**0.0179**	**0.0185**	**0.0175**	**0.0177**
	σ^2	*−0.0004*	*0.0008*	*−0.0016*	*−0.0019*	*0.0010*
		0.0407	**0.0411**	**0.0451**	**0.0406**	**0.0408**
Ignorable	μ	*0.0020*	*0.0031*	*0.0033*	*0.0035*	*0.0008*
		0.0174	**0.0296**	**0.0308**	**0.0222**	**0.0237**
	ϕ	*−0.0034*	*−0.0033*	*−0.0042*	*−0.0031*	*−0.0031*
		0.0165	**0.0215**	**0.0277**	**0.0164**	**0.0163**
	σ^2	*−0.0036*	*−0.0044*	*−0.0044*	*−0.0037*	*−0.0034*
		0.0404	**0.0426**	**0.0754**	**0.0405**	**0.0405**

4. The relative root-mean-square error of the UWML estimator of μ has the largest value, while the relative root-mean-square error of the PWML estimators of ϕ and σ^2 have the largest value, under the two informative sampling schemes in most of the cases.
5. When the sampling design is ignorable, the behaviour of the UWML, SMLL and SMLE methods of estimation is similar, in the sense that their RB's and RRMSE's have approximately the same values. The interpretation of this is that, when the sampling design is ignorable the sample and population models are the same. Also the WSML is better than PWML, because the RB and RRMSE of the estimators obtained using WSML are less than the corresponding PWML. Since we are not sure whether the design is ignorable, it is safer to use the model-based estimators, based on the sample distribution.
6. If the sampling scheme is linear or exponential we can apply the exponential model in order to estimate the parameters of the population model and incur only small losses.

From the above, we conclude that there is a loss in efficiency by using the unweighted method of estimation when the sample data has been drawn using an informative sampling scheme provided that the models are identified correctly. Also we see that, under the two sampling schemes—linear and exponential, there is no one method for dealing with informative sampling that stands out. In addition to this, if the assumed models are correct, when the sampling design is ignorable there is a considerable loss of efficiency if WSML or PWML

Table 2
RRMSE of estimators under two estimation methods

Parameter	CUWML-exponential	CWML-exponential	CUWML-linear	CWML-linear	CUWML-ignorable	CWML-ignorable
μ	0.0428	0.0444	0.0409	0.0439	0.0429	0.0733
ϕ	0.0234	0.0247	0.0175	0.0256	0.0244	0.0426
σ^2	0.0467	0.0494	0.0442	0.0483	0.0450	0.0837

First time response variable treated fixed ($T = 3$ time periods).

estimation methods are applied, while there is no considerable loss of efficiency if SMLL and SMLE are applied. This is because under the ignorable sampling scheme the sample and population distributions coincide. Also one can see that SMLE and SMLL are very robust with respect to model assumptions and in fact there is no real difference between them.

Table 2 contains the RRMSE of estimators under three estimation methods when the first time response variable is treated as fixed.

The results of Table 2 show that when the outcome of the first time period is treated as fixed, the CUWML and CWML estimators have small biases. Also the (conditional) RRMSE of the CWML estimators is larger than that of the CUWML estimators. Here we compare weighted and unweighted estimators, even though we do not need to weight, because when the response variable for the first time period is treated as fixed, then there is no problem of informativeness and the conditional sample and population distributions are the same.

Moreover, from the comparison of Tables 1 and 2, we see that the CUWML estimators have higher RRMSE than SMLL and SMLE for both of the informative sampling schemes—exponential and linear—and also for the noninformative case. This is because the latter make use of the first observation and by conditioning on the outcome of the first time period we loose some information. Thus the RRMSE of the SMLE estimator of μ is more than twice that of CWML, where the first observation is not used, so that large gains are achieved for the estimator of μ, by using the first observation. For the parameter ϕ the gains are more modest (an RRMSE 0.0165 for SMLE and 0.0247 for CWML, under exponential sampling) and for the parameter σ^2 the gains are small. However, by taking into account the first observation through informative sampling methods we achieve significant gains, especially for the mean μ but also for the autoregressive parameter ϕ.

From a comparison of results for $T = 3$ (above) to those for $T = 10$ (not given here for lack of space) we find that the relative biases and relative root-mean-square errors of the estimators of μ, ϕ and σ^2 decrease as the number of time periods increases, under all three sampling schemes. This is because the sample size increases over time. Also the effect of informativeness decreases as the length of the overall time period increases. The interpretation of this phenomenon is that since the sample is a panel sample selected at the first time period and all units remain in the sample till the last time period, the change in the superpopulation parameters happens only for the response variable at the first time period. Therefore as time increases the effect of the sample distribution at the first time

Table 3
Average values (AV) and standard deviations (SD) of K–L information statistic for testing sampling ignorability ($T = 3$ time period)

Sample model	Exponential		Linear		Ignorable	
	AV	SD	AV	SD	AV	SD
$T = 3$	14.07	2.45	10.28	1.89	0.74	1.05

period decreases. This is because as time increases the sum of the likelihood functions for $t \geqslant 2$ dominates the sum of the likelihood function for $t = 1$. Thus as the number of time points T increases the RRMSE of the estimators will converge to the same value. So the estimators SMLL and SMLE can have an advantage only when T is small.

6.6. *Tests of informativeness*

Values of the conventional t-statistic for testing the slope of the regression of w_i on y_{i1}^k is equal to zero were generated by the *cor.test* function within S-Plus (Statistical Sciences, 1990). The average values of the correlation t-test statistic for $k = 1, 2, 3$ and $T = 3, 10$, (not given here for lack of space) show that, under both the two informative sampling schemes—exponential and linear—the sampling design is highly significant, that is informative. However, under an ignorable sampling design the value of the t-statistic is very small so that the test is not significant, as it should be. The proportion of significant results as obtained for the correlation test statistics under the exponential, linear and ignorable sampling designs was calculated for five critical values $\alpha = 0.15, 0.10, 0.05, 0.025, 0.01$ of the corresponding distribution of the test statistic under the null hypothesis, which is approximately normal. The results showed that the proportions of significant results under exponential and linear sampling designs are very close to one and are close to the nominal significance levels for the ignorable design, so that the correlation test for testing sampling ignorability performs well.

Table 3 shows the average values and standard deviations of the K–L information statistic given by (26) under informative and noninformative sampling designs and under the true (exponential) and under the wrong model (linear) of sample inclusion probabilities. Here the critical value is 3.84.

From this table we conclude that the results are statistically significant under informative sampling design, when in fact the data are simulated using informative sampling designs.

Similarly, the proportion of significant results as obtained for the K–L information test statistics under the exponential, linear and ignorable sampling designs, (not given here for lack of space), was calculated for five critical values of the corresponding distribution of the test statistic under the null hypothesis, which is chi-square with one degree of freedom. The results show that the K–L information test statistic for testing sampling ignorability performs well both under informative sampling designs and under ignorable sampling design. Thus, the proportion of significant results is close to one under informative sampling design and small (and in most cases less than the nominal levels) under an ignorable sampling design.

Also the K–L information statistics is not very sensitive to wrong assumptions about the conditional expectation of sample selection probabilities.

Acknowledgements

The authors are grateful to Ray Chambers, to Danny Pfeffermann, to the Associate Editor and to the referees for their valuable comments.

References

Akaike, H., 1973. Maximum likelihood identification of Gaussian autoregressive moving average models. Biometrika 60, pp. 255–265.

Binder, D.A., 1983. On the variances of asymptotically normal estimators from complex surveys. Internat. Statist. Rev. 51, 279–292.

Binder, D.A., 1998. Longitudinal surveys: why are these surveys different from all other surveys? Survey Methodol. 24, 101–108.

Binder, D.A., Hidiroglou, M.A., 1988. Sampling in time. In: Krishnaiah, P.R., Rao, C.R. (Eds.), Handbook of Statistics, Vol. 6. Elsevier, Amsterdam, pp. 187–211.

Diggle, P.J., Kenward, M.G., 1994. Informative dropout in longitudinal data analysis (with discussion). Appl. Statist. 43, 49–93.

Diggle, P.J., Liang, K.Y., Zeger, S.L., 1994. Analysis of Longitudinal Data. Science Publication, Oxford.

Duncan, G.J., Kalton, G., 1987. Issues of design and analysis of surveys across time. Internat. Statist. Rev. 55, 97 –117.

Feder, M.F., Nathan, G., Pfeffermann, D., 2000. Time series multilevel modeling of complex survey longitudinal data. Survey Methodol. 26, 53–65.

Fienberg, S.E., 1989. Modeling considerations: discussion from a modeling perspective. In: Kasprzyk, D., Duncan, G.J., Kalton, G., Singh, M.P. (Eds.), Panel Surveys. Wiley, New York, pp. 566–574.

Friedlander, D., Okun, B.S., Eisenbach, Z., Elmakias, L.L., 2002. Immigration, social change and assimilation: educational attainment among Jewish ethnic groups in Israel. Research Reports, Israel Central Bureau of Statistics, Jerusalem.

Fuller, W., 1996. Introduction to Statistical Time Series. second ed. Wiley, New York.

Harvey, A.C., 1993. Time Series Models. second ed. Harvester Wheatsheaf, New York.

Herriot, R.A., Kasprzyk, D.,1984. The Survey of Income and Program Participation. American Statistical Association Proceedings of the Social Statistics Section, pp. 107–116.

Hoem, J.M., 1989. The issue of weights in panel surveys of individual behaviors. In: Kasprzyk, D., Duncan, G.J., Kalton, G., Singh, M.P. (Eds.), Panel Surveys. Wiley, New York, pp. 539–565.

Kasprzyk, D., Duncan, G.J., Kalton, G., Singh, M.P. (Eds.), 1989. Panel Surveys. Wiley, New York.

Kullback, S., 1978. Information Theory and Statistics. Peter Smith, Gloucester, MA.

Lawless, J.F., 2003. Event history analysis and longitudinal surveys. In: Skinner, C., Chambers, R. (Eds.), Analysis of Survey Data. Wiley, New York, pp. 221–243.

Miller, M.E., et al., 2001. A marginal model for analyzing discrete outcomes from longitudinal surveys with outcomes subject to multiple-cause nonresponse. J. Amer. Statist. Assoc. 96, 844–857.

Nathan, G., 1999. A review of sample attrition and representativeness in three longitudinal surveys. GSS Methodology Series No. 13. Office of National Statistics, London.

Pfeffermann, D., 1993. The role of sampling weight when modeling survey data. Internat. Statist. Rev. 61, 317–337.

Pfeffermann, D., 1996. The use of sampling weights for survey data analysis. Statist. Methods Med. Res. 5, 239–261.

Pfeffermann, D., Nathan, G., 2001. Imputation for wave nonresponse—existing methods and a time series approach. In: Groves, R., Dillman, D., Eltinge, J., Little, R. (Eds.), Survey Nonresponse. Wiley, New York, pp. 417–429 (Chapter 28).

Pfeffermann, D., Sverchkov, M., 1999. Parametric and semi-parametric estimation of regression models fitted to survey data. Sankhya 61 (Ser. B, Part 1), 166–186.
Pfeffermann, D., Sverchkov, M., 2003. Fitting generalized linear models under informative probability sampling. In: Skinner, C., Chambers, R. (Eds.), Analysis of Survey Data. Wiley, New York, pp. 175–195.
Pfeffermann, D., Krieger, A.M., Rinott, Y., 1998. Parametric distributions of complex survey data under informative probability sampling. Statist. Sinica 8, 1087–1114.
Skinner, C.J., 1989. Domain means, regression and multivariate analysis. In: Skinner, C.J., Holt, D., Smith, T.M.F. (Eds.), Analysis of Complex Surveys. Wiley, New York, pp. 59–88.
Skinner, C.J., 1994. Sample models and weights. American Statistical Association Proceedings of the Section on Survey Research Methods, pp. 133–142.
Skinner, C.J., Holmes, D., 2003. Random effects models for longitudinal data. In: Skinner, C., Chambers, R. (Eds.), Analysis of Survey Data. Wiley, New York, pp. 205–219.
Skinner, C.J., Holt, D., Smith, T.M.F. (Eds.), 1989. Analysis of Complex Surveys, Wiley, New York.
Smith, J.M.F., 1978. Principles and problems in the analysis of repeated survey. In: Nawboodiri, N.K. (Ed.), Survey Sampling and Measurement. New York, Academic Press, pp. 201–216.
Statistical Sciences, 1990. S-Plus Reference Manual. Statistical Sciences, Seattle.
Zeger, S.L., Liang, K.-Y., Albert, P.S., 1988. Models for longitudinal data: a generalized estimating equation approach. Biometrics 44, 1049–1060.